# Smartphone-assisted mobility in urban environments for visually impaired users through computer vision and sensor fusion

Dragan Ahmetovic
Advisor: Sergio Mascetti
Dipartimento di Informatica, Università degli Studi di Milano
dragan.ahmetovic@unimi.it

***Abstract*—For visually impaired users one of the major challenges is unassisted orientation and way-finding, especially in unexplored and potentially dangerous environments. The following work analyzes the issues stemmed from this problem and summarizes the merits and flaws of solutions available in literature. Afterwards, the research methodology is briefly described and the already achieved results are listed. Finally, a roadmap for the future contributions is proposed.**

## I. Introduction

Modern smartphones are versatile platforms, capable of ubiquitous connectivity and equipped with motion and position detection sensors. Furthermore, the computational capabilities of newer mobile phones, having dual core, fast processors (frequencies over $1000mhz$) and onboard GPUs, can be compared to modern personal computers. For users with mobility and sensory impairments, the new functionalities of these devices can be used to tackle issues caused by disabilities and lead towards a greater independence without the need of specifically designed and often expensive devices. The independent mobility of visually challenged users in unknown environments is an issue that can be effectively mitigated through the usage of video camera-based object recognition in concert with data from sensors such as accelerometers, gyroscopes and GPS.

## II. Problem domain

Mobility in urban environments can be considered as an articulated set of smaller tasks, basic ones being: navigation planning, long distance orientation and routing, sign recognition and obstacle avoidance.
Navigation planning is a simple task for a sighted user: the path, landmarks and other important information are easily conveyed through graphical elements on a map, but hardly accessible by a visually impaired user. Additionally, different information may be interesting for a blind user. For example, cues about the presence of specific hazards such as road conditions or barriers are unimportant for a sighted user but essential for a visually challenged one.
Long distance orientation can also be an issue due to the difficulty of a visually impaired user to define the direction of a distant point of interest and keep the orientation towards it without relying on haptic or tactile guides.
Sign recognition is a broad set of tasks including crosswalk, traffic lights and signs detection and identification of nearby and far landmarks. The visual aspect of these tasks makes them impossible for unassisted visually challenged users.
Short distance obstacle avoidance deals with unforeseen issues that require close range rerouting, such as traffic jams or works on roads. For blind users these trivial tasks may result complex or even dangerous without reliable assistance.
The goal of this work is to deal with each one of these tasks and problems that arise from them and ultimately design and propose a coherent smartphone-based urban navigation system for visually impaired users.

## III. Related work

As suggested in the previous section, the evolution of the human-computer interaction on mobile devices from a tactile approach (keyboards) towards a purely graphical one has led to accessibility issues for visually impaired users (McGookin et al. [9]). The touchscreen interface exploration with screen readers (such as Voiceover on iOS), which vocally describe the elements selected on the interface, certainly mitigate the problem. However, this kind of interaction may be slow since the user has to explore the interface sequentially.
The screen reader approach, aided by vibration output, can also be applied to maps (Poppinga et al. [11]). It is, however, particularly unfeasible in case of dense or overlapping elements. Heuten et al. [7] investigate the "Sonification" approach to the map exploration. The user navigates through a virtual 3D environment in which distinct audio cues corresponding to points of interest are played stereophonically in order to convey the distance and direction information. This approach requires the complete isolation of the user from the external sounds which is hard to achieve on a smartphone during mobility.
Looktel (Sudol et al. [12]) is an application suite designed to assist visually impaired users with object recognition and navigation tasks. The navigation component, called Breadcrumbs, offers GPS navigation, geotagging, path creation and social sharing capabilities. The application, however does not deal with navigation planning or Micronavigation issues.
Fallah [5] proposes a dead reckoning-based indoors navigation relying on planimetries coupled with accelerometer and compass data to estimate the user's position in absence of a GPS signal. This approach requires a detailed map of the surroundings and the estimated position error, given the imprecision of accelerometers, accrues in short time.
Pai et al. [10] suggest the usage of a pedometer model to limit the position degradation during dead reckoning by filtering out

978-0-7695-4973-6/13 $26.00 © 2013 IEEE
DOI 10.1109/MDM.2013.58

movements not corresponding to steps.
Angin et al. [3] consider a cloud computing based traffic lights recognition. The solution suffers from network delays, especially when a Wi-Fi connection is not available.
Computer vision-based detection of zebra crossings on smartphones is proposed by Ivanchenko et al. [8]. In this proposal, the position of the crosswalk, useful for the alignment of the user to the crossing, is not considered.
Computer vision-based object detection is a computation intensive, energy consuming task. The energy consumption is a concern for devices with limited battery life such as smartphones. Energy consumption patterns for hardware components on such devices have been analyzed by Carroll et al. [4].

## IV. Methodological Approach

Methodologically, the independent mobility of visually impaired users in urban contexts can be considered as two distinct problems: The first one, called "Macronavigation", deals with navigation planning and coarse origin to destination routing while the second one, "Micronavigation" deals with spatially limited issues such as sign recognition or obstacle avoidance that, while not influencing the overall route, are frequently dealt with during each step of the long range navigation.
Both problems are composed by mostly independent tasks which can be considered one at a time. The research will focus on a series of brief development cycles which will tackle each one of these issues separately.
Each development cycle will begin with an initial problem identification stage, in which a navigation issue will be singled out. The analysis of the related literature and of the problem domain will follow and the considered problem will be formalized. A suitable solution to the problem will be designed, implemented and tested while constantly interacting with the visually impaired user base in order to obtain feedback on the efficiency and usability of the proposed solutions. Issues that may arise during the development will be tackled through following iterations of the cycle.
While both Micronavigation and Macronavigation will be tackled during the course of this research, Micronavigation issues will be considered first. These tasks are less covered in the literature than the Macronavigation ones which, aside from the specialized interaction techniques involved, can easily relate to the widespread GPS-based navigation. Additionally, Micronavigation tasks are more appealing for the research since they may also benefit from the advanced sensor capabilities of modern smartphones.
Clearly, Micronavigation problems, while mostly visual in nature, also depend greatly on the orientation of the device. Thus, these problems will be dealt with by a computer vision object recognition approach backed up by sensor fusion data analysis. The output of the computed data has to consider both the special needs of the users and the environmental conditions, thus specifically designed interaction paradigms are desired. Finally, a thorough testing is needed in order to evaluate the effectiveness of the developed solutions.
Each research area involved requires specific methodologies and are considered separately in the following subsections.

### *Computer Vision*

Micronavigation problems involving the recognition of landmarks or signs will mostly be tackled through computer vision techniques. The regularity of many considered patterns allows the usage of a manually selected and tuned set of features for the detection. For example in case of zebra crossings, the pattern is a geometrically regular set of alternating light and dark stripes, having the same thickness and width (Figure 1). For less regular patterns, a machine learning approach (e.g.: adaBoosting [13]) is more suitable. Machine learning techniques use automatized classifiers along with a training set containing previously labeled data in order to infer the classification for other, unlabeled data. The lack of a small set of well defining features, as in traffic signs, shop lights and logos, makes the usage of these techniques preferred.

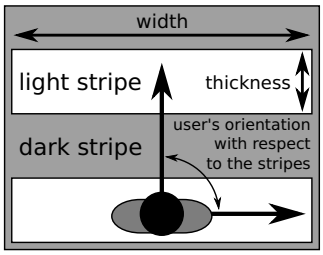


Fig. 1. Geometrical properties of zebra crossings

Computer vision techniques often require the execution of pixel-per-pixel operations (e.g: edge detection blurring filters) on captured images. These operations account for most of the computation time of object recognition algorithms. Recent smartphones often include a graphical processing unit (GPU), designed for parallel execution of operations on all pixels of an image, which will be considered in order to reduce the computation time of the adopted computer vision techniques.

### *Sensor Fusion*

Features that can be detected with computer vision often have a well defined position and orientation with respect to the user. For example zebra crossings lay on the ground plane and have stripes roughly perpendicular to the viewing direction of the user, while traffic lights are positioned on a pole at a height above the user. Accelerometers and gyroscopes allow to infer the viewing direction of the camera and thus reduce the search area on the captured images to the most plausible sections. This approach results in diminishing of both the computation time and false positive detection errors.
Once a feature has been recognized, the position of the detected object can be tracked through dead reckoning, a technique using the initial position of the object, spatial reasoning techniques and data from accelerometers and gyroscopes. This way the computational strain on the mobile device is reduced and the navigation is not interrupted even when obstacles cover the feature for short time or the user moves the camera.

### *Human Computer Interaction*

Several considerations need to be made for the design of interaction paradigms suitable for the usage in mobility by visually impaired users.
The adopted interfaces have to be appropriate with respect to the context of usage, the users needs and the information that has to be conveyed to the user.
For instance, the effectiveness of audio-based interaction is

limited in noisy environments, and even more so since a blind user often relies on auditive stimuli in order to avoid perils during movement. Diverting the user's attention from this task is potentially hazardous and should be avoided. Also touch screen interfaces, being coupled with a screen reader, should be used at most sporadically. Position-independent gestures on a touch screen, however, can be used without audio feedback and therefore are a possible interaction technique.
Two-handed interaction should also be avoided since the user might want to use the smartphone with one hand while carrying a white cane or the leash of a guide dog in the other one.
The amount of data that needs to be conveyed to the user has to be kept low and the output should avoid distracting the user from surroundings. At the same time the information about the surroundings must be delivered to the user in a timely manner, comprehensively and without errors. Haptic feedback coupled with short audio or vocal cues often yields adequate output without causing distraction to the user.

*Result Evaluation*

Through user-driven evaluations, both the accessibility of the adopted interfaces and the quality of the assistive features will be assessed. The tests will be conducted through quantitative and qualitative measurements of the interaction between the testing users and the system.
The quantitative evaluation will measure the outcomes of interactions between the users and the system, such as the number of users who were able to complete the given task, the time needed for the interactions, and other applicable metrics. The qualitative evaluation is based on the active feedback from the testing users about the perceived quality of the evaluated software and the user interface. Users can also give important insights into possible enhancements or future work. An example of guidelines for designing and evaluation of interfaces' usability is the NASA's Usability Toolkit [1].
Computer-driven evaluation will be used to assess the performance of computer vision and sensor fusion-based components of the system. The navigation and orientation tasks, given the safety concerns for the user, are time critical, thus sufficiently low (near real-time) computation times will be required.
Precision metric, which expresses the ratio of false positive erroneous detections, is the most delicate one due to the safety risks that faulty detections during the navigation may cause. This metric will often be required to yield perfect results (No false positives). For the recall metric, which expresses the amount of undetected features, the goal is the best possible detection, but without impacting the precision or the computation time of the solution.

## V. Achieved Results

During the initial stages, the research focused on a single object recognition problem: the zebra crossing recognition on a single captured picture. In this work the recognition consists in the detection and validation of a set of geometrical and visual features common to all the zebra crossings.
First, a hough-based segment detector (OpenCV library implementation [2]) tests the input image for the presence of long line segments. The segments are grouped according to parallelism, distance constraints between close segments and existence of alternating black and white color blocks between them. A possible group of suitable segments closest to the bottom of the image is labeled as crossing and its coarse position with respect to the user is output. Figure 2 shows an example of the output of the algorithm.
The recognition algorithm was evaluated on a test set of 400 pictures at different resolutions, half of which containing crossings. The experiments yielded a precision result of 1 (no false positives were detected), while the recall metric result on the selected test set was 0.75. The execution time of 365 milliseconds on an iPhone 3gs platform, however, was not compatible with real-time detection.

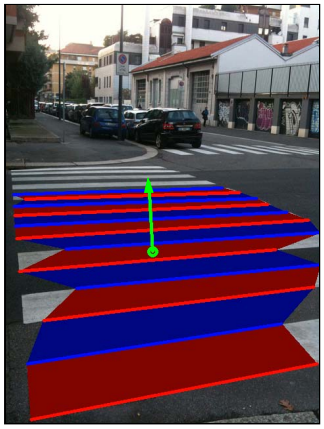

Fig. 2. Result of zebra crossing recognition algorithm

An extension of this work was proposed in [1]. In this solution the detection is executed on a video stream and the position of the smartphone is inferred through the usage of accelerometers. The accelerometer data is used to calculate the horizon inside the captured images and a modified version of the LSD segment detector algorithm (Grompone von Gioi et al. [6]), is used. The modified version only considers the areas under the horizon and only the segments roughly parallel to the horizon are detected. The extension also explores the user interaction and a novel interaction technique consisting in short vocal messages coupled with vibration feedback is proposed.
Finally, the solution is evaluated through user and computer driven evaluations. The computer-based tests show that the detection improves the previous results yielding a computation time of 46 milliseconds while maintaining a perfect precision score. The recall value of 0.65, while lower than in the previous work, is improved through result aggregation of 3 consecutive runs resulting in the correct detection of all crossings during the user tests. User-driven evaluations with 5 visually impaired users have also shown that the solution correctly guides and aligns the user with respect to the zebra crossings.

## VI. Research and Expected Contributions

The ultimate goal of this research will be the design, development and evaluation of a mobile phone-based navigation software for visually impaired users dealing with both Micronavigation and Macronavigation problems. The main components of the solution will be:

- a map viewing and route planning component

[1] http://www.hq.nasa.gov/pao/portal/usability/
[2] http://opencv.org/

- a GPS-based coarse navigation tool
- an obstacle and landmark recognition and avoidance component
- a tool for geotagging and sharing of points of interest
- audio and haptic-based interface for interaction with the previous components

Specifically, contingent research aims to improve the position calculation of previous works and, through the usage of dead reckoning techniques, guide the user towards a crossing even when obstacles temporarily occlude the view or the user moves and rotates the camera away from the zebra crossing.
As anticipated in Section II, error accumulation is an important issue of this approach. While the gyroscope data degrades over tens of seconds, the position, calculated jointly with accelerometers and gyroscopes, degrades quickly (meters of error in a second). This is caused by the double integration of the accelerometer data, required to obtain the position information from the acceleration, which transforms the noise from the original sensor readings into a consistent drift which accrues quickly. In order to reduce the accumulation of error, the use of a pedometer algorithm (Pai et al. [10]) will be used.
The existing object recognition avoids false positives and has acceptable computation times at the expense of lower recall values. A desired future contribution is the usage of GPU-based image filtering. This approach is expected to achieve better execution times and thus would allow the usage of higher resolution images, resulting, in turn, in better recall values.
The recent development of faster segment detection algorithms, such as EDlines (Akinlar et al. [2]), would reduce the number of required image filtering steps and would possibly contribute cut the execution times of the object recognition in half.
The recognition of pedestrian traffic lights is also being considered. Similarly to the zebra crossing recognition, accelerometers are used to limit the search area and guide the user correctly. The object recognition focuses on three distinct cues: the position of the light on the traffic light box, the shape of the light and the color spectrum of the light.
The position of the light in the box is a straightforward feature; if the light is red, its position is on top of the box, if it's yellow it's in the middle and if it's green it's on the bottom. The size of the light, the box and their proportion can be used to verify if the considered object is actually a traffic light and its distance.
The shape of the light can help to discriminate between car lights and pedestrian lights and output the information to the user accordingly.
The color spectrum of the light is a more involved feature. In different lighting conditions the spectrum of a traffic light can change profoundly. This notion can be used to craft specific color ranges for different illumination patterns of the scene and improve both the precision and the recall values of the detection algorithm.
As hinted in Section II, visually challenged users may be interested in information about places and routes that differ from those useful to the sighted users. Notions about the topographical conformation of a place, perils specific to the blind users or points of interests are noteworthy data for visually impaired users. Suitable cues, both inserted by the user and automatically detected, or even routes can be shared with other users through geotagging capabilities of the solution.
Currently adopted interaction techniques are mostly audio based. For orientation tasks, short audio cues, which do not distract the user, can be adopted. Complex interactions, however, still rely on the screen reader-driven touch interface. Before triggering an element of the interface, the screen has to be explored, the element selected and then activated. The usage of this interaction paradigm in noisy environments is still unfeasible. Simple vibration pattern output in place of audio cues will be evaluated. For more complex input, instead, position independent gestures on screen or by shaking and moving the whole device will be investigated.

## ACKNOWLEDGMENTS

I would like to thank my advisor dr. Sergio Mascetti, prof. Claudio Bettini, Cristian Bernareggi and Andrea Gerino for their assistance and insight during my Ph.D work.

## REFERENCES

[1] Ahmetovic, D., Bernareggi, C., and Mascetti, S. Zebralocalizer: identification and localization of pedestrian crossings. In *Proceedings of 13th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2011.

[2] Akinlar, C. and Topal, C. Edlines: Real-time line segment detection by edge drawing (ed). In *Proceedings of 18th International Conference on Image Processing (ICIP)*. IEEE, 2011.

[3] Angin, P., Bhargava, B., and Helal, S. A mobile-cloud collaborative traffic lights detector for blind navigation. In *Proceedings of International Conference on Mobile Data Management*. IEEE, 2010.

[4] Carroll, A. and Heiser, G. An analysis of power consumption in a smartphone. In *Proceedings of the 2010 USENIX annual technical conference*. USENIX Association, 2010.

[5] Fallah, N. Audionav: a mixed reality navigation system for individuals who are visually impaired. In *SIGACCESS Access. Comput.* ACM, 2010.

[6] Grompone von Gioi, R., Jakubowicz, J., Morel, J.M., and Randall, G. Lsd: A fast line segment detector with a false detection control. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[7] Heuten, W., Wichmann, D., and Boll, S. Interactive 3d sonification for the exploration of city maps. In *Proceedings of the 4th Nordic Conference on Human-computer interaction*. ACM, 2006.

[8] Ivanchenko, V., Coughlan, J., and Shen, H. Crosswatch: a camera phone system for orienting visually impaired pedestrians at traffic intersections. In *Proceedings of 11th International Conference on Computers Helping People with Special Needs (ICCHP '08)*. Springer, 2008.

[9] McGookin, D., Brewster, S., and Jiang, W. Investigating touchscreen accessibility for people with visual impairments. In *Proceedings of Nordic Conference on Human Computer Interaction (NordiCHI)*. ACM, 2008.

[10] Pai, D., Malpani, M., Sasi, I., Aggarwal, N., and Mantripragada, P.S. Padati: A robust pedestrian dead reckoning system on smartphones. In *Proceedings of 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*. IEEE, 2012.

[11] Poppinga, B., Pielot, M., Magnusson, C., and Rassmus-Grhn, K. Touchover map: Audio-tactile exploration of interactive maps. In *Proceedings of 13th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, 2011.

[12] Sudol, J., Dialameh, O., Blanchard, C., and Dorcey, T. Looktel, a comprehensive platform for computer-aided visual assistance. In *Proceedings of Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, 2010.

[13] Viola, P. and Jones, M. Robust real-time face detection. In *Internationall Journal of Computer Vision (IJCV)*. Springer, 2004.